RAW SEQUENCE LISTING
ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: _10/009,792_
Source: _PCT10_
Date Processed by STIC: _7/11/02_

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:
1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY
FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.
PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)
PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:
    http://www.uspto.gov/web/offices/pac/checker

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.
Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.
Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<http://www.uspto.gov/ebc/efs/downloads/documents.htm> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:
   U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7th Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202
        Or
   U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
        Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 01/29/2002

PCT10

Doe

RAW SEQUENCE LISTING   DATE: 07/11/2002
PATENT APPLICATION: US/10/009,792   TIME: 14:06:50

Input Set : A:\seqlist.txt
Output Set: N:\CRF3\07112002\J009792.raw

```
 3 <110> APPLICANT: Korea Advanced Institute of Science and Technology
 5 <120> TITLE OF INVENTION: E coli Secreting Human
 6       Granulocyte Colony Stimulating Factor(hG-CSF)
 8 <130> FILE REFERENCE: P0132/KAIST/PCT
C--> 10 <140> CURRENT APPLICATION NUMBER: US/10/009,792
C--> 10 <141> CURRENT FILING DATE: 2001-12-13
   10 <150> PRIOR APPLICATION NUMBER: KR 2000-0017052
   11 <151> PRIOR FILING DATE: 2000-03-31
   13 <160> NUMBER OF SEQ ID NOS: 27
   15 <170> SOFTWARE: KopatentIn 1.71
```

ERRORED SEQUENCES

```
   280 <210> SEQ ID NO: 19
   281 <211> LENGTH: 174
   282 <212> TYPE: PRT
   283 <213> ORGANISM: Homo sapiens
   285 <400> SEQUENCE: 19
   286 Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys
E--> 289 Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu Gln
E--> 292 Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu Val
E--> 295 Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser Cys
E--> 298 Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His Ser
E--> 301 Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile Ser
E--> 304 Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala Asp
E--> 307 Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala Pro
E--> 310 Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala Phe
E--> 313 Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser Phe
E--> 316 Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro
   345 <210> SEQ ID NO: 21
   346 <211> LENGTH: 175
   347 <212> TYPE: PRT
   348 <213> ORGANISM: Homo sapiens
   350 <400> SEQUENCE: 21
   351 Met Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu
E--> 354 Lys Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala Leu
E--> 357 Gln Glu Lys Leu Cys Ala Thr Tyr Lys Leu Cys His Pro Glu Glu Leu
E--> 360 Val Leu Leu Gly His Ser Leu Gly Ile Pro Trp Ala Pro Leu Ser Ser
E--> 363 Cys Pro Ser Gln Ala Leu Gln Leu Ala Gly Cys Leu Ser Gln Leu His
E--> 366 Ser Gly Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Gly Ile
E--> 369 Ser Pro Glu Leu Gly Pro Thr Leu Asp Thr Leu Gln Leu Asp Val Ala
E--> 372 Asp Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly Met Ala
```

(handwritten: numbering underneath every 5th amino is required. See Error Summary sheet item 3)

```
          RAW SEQUENCE LISTING              DATE:  07/11/2002
          PATENT APPLICATION:  US/10/009,792    TIME:  14:06:50

          Input Set  : A:\seqlist.txt
          Output Set: N:\CRF3\07112002\J009792.raw


E--> 375 Pro Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe Ala Ser Ala
E--> 378 Phe Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser His Leu Gln Ser
E--> 381 Phe Leu Glu Val Ser Tyr Arg Val Leu Arg His Leu Ala Gln Pro
     417 <210> SEQ ID NO: 25
     418 <211> LENGTH: 45
     419 <212> TYPE: PRT
     420 <213> ORGANISM: Homo sapiens
     422 <400> SEQUENCE: 25
     423 Met Phe Lys Phe Lys Lys Lys Phe Leu Val Gly Leu Thr Ala Ala Phe
E--> 426 Met Ser Ile Ser Met Phe Ser Ala Thr Ala Ser Ala Thr Pro Leu Gly
E--> 429 Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu Lys Cys
     446 <210> SEQ ID NO: 27
     447 <211> LENGTH: 60
     448 <212> TYPE: PRT
     449 <213> ORGANISM: Homo sapiens
     451 <400> SEQUENCE: 27
     452 Met Phe Lys Phe Lys Lys Lys Phe Leu Val Gly Leu Thr Ala Ala Phe
E--> 455 Met Ser Ile Ser Met Phe Ser Ala Thr Ala Ser Ala Ala Gly Pro His
E--> 458 His His His His His Ile Glu Gly Arg Thr Pro Leu Gly Pro Ala Ser
E--> 461 Ser Leu Pro Gln Ser Phe Leu Leu Lys Cys Leu Glu
E--> 469 1
E--> 472 6   delete
```

VERIFICATION SUMMARY   DATE: 07/11/2002
PATENT APPLICATION: US/10/009,792   TIME: 14:06:51

Input Set : **A:\seqlist.txt**
Output Set: **N:\CRF3\07112002\J009792.raw**

```
L:10 M:270 C: Current Application Number differs, Replaced Current Application No
L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:289 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:19
M:332 Repeated in SeqNo=19
L:354 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:21
M:332 Repeated in SeqNo=21
L:426 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:25
M:332 Repeated in SeqNo=25
L:455 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:27
M:332 Repeated in SeqNo=27
```

Raw Sequence Listing Error Summary

ERROR DETECTED	SUGGESTED CORRECTION	SERIAL NUMBER: 10/009,792

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

1 ____ Wrapped Nucleics
Wrapped Aminos

The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."

2 ____ Invalid Line Length

The rules require that a line not exceed 72 characters in length. This includes white spaces.

3 ✓ Misaligned Amino Numbering

The numbering under each 5th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.

4 ____ Non-ASCII

The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.

5 ____ Variable Length

Sequence(s) ____ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.

6 ____ PatentIn 2.0 "bug"

A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) ____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.

7 ____ Skipped Sequences (OLD RULES)

Sequence(s) ____ missing. If intentional, please insert the following lines for each skipped sequence:
(2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
(i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)
(xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)
This sequence is intentionally skipped

Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.

8 ____ Skipped Sequences (NEW RULES)

Sequence(s) ____ missing. If intentional, please insert the following lines for each skipped sequence.
<210> sequence id number
<400> sequence id number
000

9 ____ Use of n's or Xaa's (NEW RULES)

Use of n's and/or Xaa's have been detected in the Sequence Listing.
Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.

10 ____ Invalid <213> Response

Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence

11 ____ Use of <220>

Sequence(s) ____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.
(See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)

12 ____ PatentIn 2.0 "bug"

Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

13 ____ Misuse of n

n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.

AMC/MH – Biotechnology Systems Branch – 08/21/2001